

Grzegorz Bryda

Magdalena Jelonek

Barbara Worek

Jak skonstruować dobre narzędzie do ewaluacji jakości zajęć dydaktycznych? Refleksje praktyczne

Niniejszy artykuł poświęcony został praktyce ewaluacji jakości zajęć dydaktycznych, w tym możliwościom w zakresie wykorzystania ilościowych i jakościowych metod oceny pracy nauczycieli akademickich. Zaprezentowane zostały w nim takie kwestie, jak:

- etapy konstruowania narzędzia do oceny jakości kształcenia w szkole wyższej;
- podstawowe zasady konstrukcji pytań oceniających pracę nauczycieli akademickich;
- czynniki, które mogą zaburzyć rzetelną ocenę jakości dydaktyki oraz możliwości w zakresie zapobiegania wpływowi tych czynników na wyniki badań;
- najczęściej popełniane błędy w procesie konstruowania narzędzi ewaluacyjnych, zasady analizy oraz prezentacji materiału empirycznego.

Artykuł oprócz opracowania teoretycznego zawiera przykłady praktyczne oparte zarówno na praktykach międzynarodowych, jak i na rozwiązaniach, które wdrożył Uniwersytet Jagielloński.

Wprowadzenie

Wydaje się, że ocenianie zajęć dydaktycznych przez studentów przestaje być powoli w polskim systemie szkolnictwa wyższego czymś unikatowym, świadczącym o tym, że uczelnia przywiązuje wagę do jakości kształcenia oraz że traktuje studentów w sposób partnerski, doceniając znaczenie ich opinii. Ponieważ posiadanie wewnętrznego systemu zapewniania jakości kształcenia, obejmującego też badania opinii studenckich, jest warunkiem uzyskania pozytywnej oceny w procesie akredytacji, każda uczelnia starająca się o nią musi wykazać się posiadaniem takiego systemu. Formalnie można więc przyjąć, że wszystkie uczelnie posiadające akredytację do prowadzenia studiów danego stopnia na konkretnym kierunku, podejmują działania zmierzające do oceny jakości kształcenia i uwzględniają w tych działaniach opinie studenckie jako źródło danych na temat przebiegu procesu dydaktycznego¹.

Doceniając zatem fakt, że prowadzenie badań oceniających jakość dydaktyki staje się powoli obowiązującym standardem, chcemy zauważyć, że badania te będą wartościowym i użytecznym źródłem informacji tylko wtedy, gdy będą prowadzone w sposób gwarantujący trafność i rzetelność wyników oraz gdy ich realizacji będzie towarzyszyła refleksja dotycząca celu i sposobu wykorzystania zebranych danych. W tym przypadku, podobnie jak i w wielu innych, kluczowy jest bowiem nie tyle sam fakt badania opinii studenckich, ile to, jak owe badania są prowadzone oraz jak wykorzystywane są ich wyniki.

1 Warto tutaj zaznaczyć, że samo pojęcie „jakość kształcenia” jest różnie definiowane i rozumiane. Jakość może być rozumiana jako ceniony zbiór cech procesu dydaktycznego wobec stanu rzeczywistego lub możliwego do osiągnięcia czy, też jako stabilność założonych parametrów przyjętego modelu kształcenia. Jakość kształcenia jest postrzegana z dwóch perspektyw: wewnętrznej i zewnętrznej (Payne 1996:264). Pierwsza, wskazuje na zgodność procesu kształcenia z określonymi normami (np. standardy akredytacyjne), druga, na subiektywną ocenę doświadczeń jednostek uczestniczących w procesie kształcenia (np. ocena zajęć dydaktycznych przez studentów). Definicje jakości kształcenia można sprowadzić do dualnego podziału, zgodnie z którym jakość to zgodność z przyjętymi standardami (kryteria obiektywne) lub jakość to interpretowana użyteczność jednostkowa (kryteria subiektywne). W praktyce trudno utrzymać ten podział. Obecnie łączy się te dwa aspekty jakości z wyraźnym naciskiem na użyteczność – kształcenie umiejętności. Świadczą o tym np. dyskusje dotyczące zwiększenia autonomii programowej szkół wyższych i wypracowania sposobów ewaluacji efektów procesu kształcenia.

Na trafność² i rzetelność³ badań wpływa wiele czynników, które nie sposób omówić w tak krótkim opracowaniu. Upraszczając nieco problem, można przyjąć, że wpływa na nie poprawna conceptualizacja badań, ich właściwa operacjonalizacja, przeprowadzenie zgodnie z przyjętymi dla danej techniki standardami, poprawna analiza danych, uwzględniająca wpływ potencjalnych źródeł obciążeń wyników⁴. Omówienie wszystkich tych kwestii wymagałoby obszernego opracowania, dlatego w niniejszym artykule skoncentrujemy się przede wszystkim na narzędziu badawczym i możliwych błędach, jakie mogą się pojawiać przy jego konstrukcji. Budowę narzędzia badawczego będziemy jednak rozpatrywać w powiązaniu z innymi etapami realizacji badań: odniesiemy się do niezbędnych działań, które powinny poprzedzać konstrukcję kwestionariusza do badań opinii studenckich na temat jakości zajęć, omówimy czynniki, które mogą obciążać wyniki takich badań, przedstawimy też podstawowe zasady analizy ich wyników i sposoby prezentacji danych. Na koniec odniesiemy się do różnego rodzaju obaw związanych z oceną zajęć, występujących zarówno wśród nauczycieli akademickich, jak i wśród studentów.

Etapy konstruowania narzędzia do oceny jakości kształcenia w szkole wyższej

Posiadanie trafnego i rzetelnego narzędzia badawczego, które pozwoli monitorować jakość usług edukacyjnych, jest kluczową potrzebą każdego systemu oceny jakości kształcenia w szkole wyższej. Najczęściej narzędziem tym jest kwestionariusz ankiety ewaluacyjnej, dystrybuowany w formie papierowej lub elektronicznej (online), np. w systemie USOS⁵. Kwestionariusz ankiety ewaluacyjnej musi wynikać z logiki zastosowanego modelu ewa-

2 Cecha narzędzia, która wskazuje na to, że mierzy ono to, co w intencji twórców powinno mierzyć.

3 Cecha narzędzia, która określa dokładność pomiaru.

4 Wymienione czynniki odpowiadają przede wszystkim etapom i działaniom związanym z realizacją badań ilościowych, prowadzonych technikami sondażowymi. Pomijamy tutaj kwestie związane z trafnością i wiarygodnością danych zbieranych za pomocą innych technik oceny zajęć dydaktycznych, choć techniki te mogą, a nawet powinny być wykorzystywane jako komplementarne w stosunku do technik sondażowych. Ponieważ badania ilościowe są jednak wykorzystywane najpowszechniej, zdecydowaliśmy się do nich zawęzić analizę.

5 Niektóre jednostki akademickie stosują uproszczone, wewnętrzne systemy prowadzenia badań ewaluacyjnych, oparte na innych niż USOS rozwiązaniach. Najczęściej sytuacja ta dotyczy szkół prywatnych.

luacji i strategii metodologicznej przyjętej do oceny procesu jakości kształcenia w szkole wyższej. Przykładowo, na Uniwersytecie Jagiellońskim zastosowanie do oceny zajęć dydaktycznych systemu USOS (ankiety w wersji online) spowodowało systematyczny wzrost liczby zajęć i pracowników objętych oceną⁶. Zważywszy na to, że system jest względnie nowy, przyjęta strategia ma na celu przede wszystkim „pokrycie oceną” w formie ankiety elektronicznej wszystkich zajęć i pracowników uczelni, zaś dopiero w drugiej kolejności zwiększanie frekwencji oceniających studentów.

W praktyce problem konstruowania narzędzia ewaluacji zajęć sprowadza się do wyboru jednej z dwóch strategii działania. Pierwsza to zbudowanie od podstaw modelu jakości kształcenia poprzez wyróżnienie kluczowych wymiarów pojęcia, a następnie skonstruowanie modelu pomiarowego (systemu wskaźników) do każdego z jego podwymiarów. Druga opiera się na rekonstrukcji modelu jakości kształcenia w oparciu o istniejące już rozwiązania cząstkowe (poprawa systemu, który już funkcjonuje w uczelni wyższej). Jeszcze inne rozwiązanie to kopiowanie wzorców – narzędzi ewaluacji – europejskich lub polskich szkół wyższych (tzw. izomorfizm mimetyczny).

Oczywiście zanim przystąpimy do budowy własnego narzędzia – kwestionariusza – należy sobie zadać pytanie, czy kwestionariusz ankiety jest rzeczywiście najlepszym narzędziem umożliwiającym odpowiedź na pytania ewaluacyjne? Czy inne metody nie są bardziej adekwatne do oceny jakości kształcenia? Niezależnie od tego, modelowa ścieżka konstrukcji narzędzia ewaluacji w formie kwestionariusza ankiety wymaga wykorzystania w pierwszej fazie metod jakościowych, najczęściej w postaci serii grup fokusowych (ewentualnie wywiadów pogłębionych) wśród studentów i pracowników. Pomagają one w rekonstrukcji wymiarów oceny jakości kształcenia, poprzez zidentyfikowanie elementów procesu kształcenia, które zdaniem badanych są najważniejsze. W niektórych przypadkach wywiady grupowe pomóc mogą w operacjonalizacji (m.in. uzupełnienie kafeterii odpowiedzi, pierwszy test narzędzia). Niezwykle ważnym, a jednak często pomijanym etapem w procesie konstrukcji kwestionariusza oceny studenckiej, jest pilotaż narzędzia. Pilotaż, będący drugim po preteście testem kwestionariusza, dostarcza racjonalnych podstaw ku temu, żeby wyselekcjonować

⁶ W roku akademickim 2006/2007 ocenionych w formie elektronicznej było 67% wszystkich zajęć/86% pracowników, w roku 2007/2008 – 78% zajęć/93% pracowników, zaś w roku 2008/2009 – 86% zajęć/96% pracowników.

te pozycje, które najlepiej wskazują na ukryte wymiary jakości kształcenia, a także te, które w największym stopniu różnicują badane obiekty. Tym samym umożliwia maksymalną redukcję liczby pytań w ankiecie, przy minimalnej stracie wyjaśnionej wariancji ukrytego konstruktu. Najczęściej wykorzystywanymi w tym zakresie technikami statystycznymi są: eksploracyjna analiza czynnikowa (EFA) oraz analiza głównych składowych (PCA).

Kwestionariusz ankiety ewaluacyjnej musi być krótki, lecz treściwy, co nie jest łatwe, szczególnie wtedy, gdy w ocenie elektronicznej student dostaje przeciętnie 6–10 kwestionariuszy odnoszących się do zajęć, na które uczęszcza. W praktyce – dzięki temu, że system jest otwarty przez określony czas – studenci wracają, aby wypełnić ankiety, które im pozostały z wcześniejszej oceny.

Jak pokazują doświadczenia konstruowania kwestionariusza ankiety ewaluacyjnej na Uniwersytecie Jagiellońskim, aby możliwe były późniejsze porównania między jednostkami, zajęciami czy pracownikami uczelni, niezbędna jest standaryzacja narzędzia. Jej dobrym uzupełnieniem jest „kontekstualizacja” narzędzia, uzyskiwana m.in. przez zastosowanie skal z opisanymi krańcami (tzw. dyferencjał semantyczny). W przypadku Uniwersytetu Jagiellońskiego wykorzystanie dyferencjału semantycznego dało klarowny układ odniesienia dla każdego pytania ankiety oceny zajęć dydaktycznych (podejście to jednak nadal jest w fazie testów).

Na podstawie wcześniejszych doświadczeń z oceną zajęć dydaktycznych na każdym wydziale uznano, że najlepszym rozwiązaniem z punktu widzenia budowy uniwersyteckiego systemu doskonalenia jakości kształcenia jest elektroniczna ocena zajęć dydaktycznych. Standaryzacja elektronicznej ankiety ewaluacyjnej została przeprowadzona na poziomie całego Uniwersytetu, z uwzględnieniem typów zajęć, różnic między jednostkami i przedmiotami w danej jednostce. Przy standaryzacji ankiety ewaluacyjnej kierowano się zasadą zachowania autonomii i uwzględniania specyfiki poszczególnych jednostek Uniwersytetu Jagiellońskiego (wydziałów, instytutów lub katedr). Istotnym i cenionym, szczególnie przez wykładowców, elementem oceny zajęć stała się możliwość uzupełnienia oceny ilościowej poprzez studenckie komentarze, uwagi i sugestie.

Ocena pracy nauczyciela akademickiego – zasady konstrukcji pytań kwestionariuszowych

Skonstruowanie dobrego narzędzia oceniającego jakość prowadzonych zajęć wymaga podjęcia szeregu decyzji. W pierwszej kolejności odpowiedzieć musimy na pytania koncepcyjne i operacyjne. Jakich informacji potrzebujemy?⁷ W jaki sposób te informacje będą wykorzystywane wewnątrz uczelni? W jaki sposób, w jakiej formie (wartości zagregowane/indywidualne) oraz komu prezentowane będą wyniki? W drugiej kolejności podjąć musimy decyzje metodologiczne. Jakiej metody użyć? Jaką formę pytania wybrać (pytania otwarte/zamknięte)? Jak sformułować pytania i kafeterie odpowiedzi? Czy z punktu widzenia założonej metodologii liczba pytań jest wystarczająca, a może za duża? Jak uniknąć zniechęcenia badanych? W jakiej kolejności zadać pytania?

W większości kwestionariuszy oceniających pracę nauczycieli akademickich znajdują się zarówno pytania o fakty (np. dostępność sylabusu, literatury przedmiotu), jak i o opinie czy subiektywne doświadczenia (ocena prowadzącego zajęcia, por. Brzezińska, Brzeziński 2004; Oppenheim 1992). W przypadku obu typów pytań kwestionariuszowych wyróżnić można obowiązujące schematy ich konstrukcji oraz najczęściej spotykane przy ich tworzeniu błędy. Przykładowo, pytania o fakty, które mają na celu uzyskanie obiektywnych informacji, zawsze powinny umożliwiać respondentowi wybór opcji „nie wiem”/„nie potrafię sobie przypomnieć”. W procesie oceny pracy nauczycieli akademickich zdecydowanie częściej wykorzystuje się klasyczne skale ocen szacunkowych, w których to kategorie odpowiedzi stanowią kwantyfikatory wskazujące na stopień intensywności danej oceny (por. Brzezińska, Brzeziński 2004). Przykładowo, skale

7 W związku z wielowymiarowym charakterem jakości kształcenia i nieraz odmiennym definiowaniem przez autorów składających się na to pojęcie wymiarów trudno jest wyróżnić wszystkie elementy budujące wysoką jakość nauczania. Za Williamem E. Cashinem (1995) możemy wymienić najczęściej oceniane w ankietach ewaluacyjnych obszary jakości, którymi są: sposób organizacji zajęć, jasność prezentacji i umiejętności komunikacyjne nauczyciela akademickiego, stosunek nauczyciela do studentów, stopień trudności kursu, sposób oceny pracy studenta, samoocena studenta w zakresie efektów uczenia się. Ponadto kategorie te uzupełnić można o: ocenę koncepcji i formy zajęć, atrakcyjność zajęć, przygotowanie nauczyciela do zajęć, zaangażowanie prowadzącego, umiejętność zainteresowania przedmiotem nauczania, umiejętność wykorzystania czasu czy gotowość do udzielania dodatkowej pomocy. W badaniach często pojawia się dodatkowe pytanie o skłonność do rekomendowania danego kursu innym studentom, które traktowane jest jako jeden z kluczowych wskaźników stopnia zadowolenia z uczestnictwa w kursie.

takie przybrać mogą postać klasycznego pytania typu Likerta z kategoriami odpowiedzi od „zdecydowanie się zgadzam” do „zdecydowanie się nie zgadzam”.

W związku z ograniczoną objętością niniejszego artykułu poniżej, w sposób skrótowy, przedstawiono zasady konstrukcji pytań kwestionariuszowych służących do oceny pracy nauczyciela akademickiego. Zasady te ograniczone zostały do najważniejszych reguł, a jednym z kryteriów ich wyróżnienia była częstość „łamania” danego założenia w praktyce badawczej. W związku z tym, że najłatwiej uczyć się na własnych bądź cudzych błędach, do każdej z reguł dopasowane zostały przykłady błędnie sformułowanych pytań, które łamią daną regułę. Pytania te pochodzą z kwestionariuszy ewaluacyjnych, stosowanych w praktyce, które zgromadzone zostały dla potrzeb niniejszego artykułu⁸.

Reguła 1. Zadawaj pytania w sposób jasny, jednoznaczny, zrozumiały i dopasowany do typu zajęć (np. wykład / konwersatorium / laboratorium).

W wielu analizowanych kwestionariuszach ewaluacyjnych pojawiają się pytania, które mogą być **wieloznacznie rozumiane**, jak np. prośba o ocenę stwierdzenia: „zachowane są proporcje między zajęciami i przerwami” czy o ocenę kategorii nazwanej: „poziom organizacji struktury treści wykładu”. Pytania takie mogą niejednokrotnie generować odpowiedzi różnorodnie interpretowane.

W jednym z kwestionariuszy znajdujemy tak sformułowane pytanie: „Czy zajęcia odbywały się regularnie?” O czym będzie świadczyła przewaga odpowiedzi, że nie? O tym, że prowadzący nagminnie przekładał zajęcia czy może o tym, że zajęcia odbywały się w systemie blokowym (a więc nieregularnie)?. Samo sformułowanie „regularnie” budzi wiele wątpliwości, podobnie zresztą jak kategorie odpowiedzi typu: „często”, „czasami”, „rzadko”, które mogą posiadać odmienne znaczenia dla każdego z respondentów. Błąd taki wkraść się do niżej zaprezentowanego pytania: „proszę podać częstotliwość obecności na zajęciach z ocenianego przedmiotu” (kafeteria odpowiedzi: „zawsze”, „w większości”, „sporadycznie”, „nigdy”)⁹.

8 Nie identyfikujemy w tekście szkół, z których pochodzą przykładowe pytania.

9 Alternatywne, wyżej oceniane podejście zastosowane zostało w narzędziu ewaluacyjnym innej szkoły wyższej, w którym umieszczono pytanie: „jaka była Pani/Pana frekwencja na zajęciach z tego przedmiotu?” (kafeteria odpowiedzi: 1=0–20%; 2=21–40%; 3=41–60%; 4=61–80%; 5=81–100%).

Podobny problem z niejednoznacznością może wystąpić, kiedy prosimy respondenta o **ocenę bardzo ogólnej kategorii**, jaką jest przykładowo organizacja zajęć. W związku z tym, że organizacja ta składa się z wielu elementów, takich jak punktualność lub dostęp do sylabusu, należy zadawać pytania raczej o elementy szczegółowe, a nie o ogólne.

Reguła 2. Unikaj pytań długich, skomplikowanych i napisanych „językiem branżowym”, technicznym czy żargonem.

Jako „język branżowy” należy rozumieć np. sformułowania takie jak „jakość kształcenia”. Jeżeli zatem chcemy o nią zapytać, to albo musimy ją skrótowo zdefiniować, co może okazać się jednak zbyt dużym wyzwaniem, lub potraktować jako pojęcie latentne (nieobserwowane w sposób bezpośredni) i zanalizować w oparciu o baterię pytań kwestionariuszowych (innymi słowy, zbudować skalę do jej pomiaru).

Reguła 3. Nie zadawaj pytań, na które odpowiedź jest oczywista.

Zadając pytanie: „Czy zawsze byłeś/aś skoncentrowany/a w trakcie zajęć?”, nie spodziewajmy się, że odpowiedzi studentów różnicować będą w sposób istotny zajęcia, w których uczestniczyli. Nie ma chyba osoby, która nawet na najbardziej interesujących zajęciach byłaby bez przerwy skoncentrowana.

Reguła 4. Unikaj zadawania pytań, na które nie jest możliwa jednoznaczna odpowiedź.

Częstą praktyką łamiącą powyższą regułę są **pytania o dwie rzeczy na raz**, które pojawiają się w większości kwestionariuszy ewaluujących pracę nauczycieli akademickich i wprowadzają badanego w stan zakłopotania, w momencie gdy dwa elementy zawarte w pytaniu powinny zostać ocenione odmiennie. „Czy prowadzący zajęcia był życzliwy, otwarty i taktowny w stosunku do studentów?” (jaką odpowiedź powinien zaznaczyć respondent, jeżeli prowadzący był życzliwy, ale nie otwarty?); „Czy materiały dydaktyczne do zajęć były dostępne i przydatne?” (a jeśli były dostępne, ale nieprzydatne?); „Czy ustalone terminy zajęć i konsultacji są przez prowadzącego przestrzegane?” (a jeśli przestrzegane są terminy zajęć, a konsultacji już nie?); „Czy prowadzący podał przejrzyste zasady zaliczenia zajęć na początku semestru i zastosował się do nich?” (a jeśli podał, ale się do nich nie zastosował?).

Ponadto nie można wymagać od respondenta dokonywania wyborów niemożliwych. Jeśli w jednym z pytań: „Podaj orientacyjną średnią ocen ze wszystkich przedmiotów, uzyskaną

w ostatnim semestrze” autor zaproponował następujące odpowiedzi: 2,0–3,0; 3,0–4,0; 4,0–5,0, to w związku z **nierozłącznością przedziałów** respondent, którego średnia ocen wynosi 3.0 lub 4.0 nie jest w stanie dokonać jednoznacznego wyboru.

Reguła 5. Skoncentruj się na pytaniach zamkniętych, ale nie unikaj pytań otwartych.

Mimo że główna zasada konstruowania narzędzi do badań kwestionariuszowych mówi, iż należy unikać pytań otwartych, w przypadku ewaluacji jakości kształcenia pytania te mogą dostarczyć wielu cennych informacji, takich jak: rekomendacje odnośnie zmiany formy, struktury czy treści zajęć, opinie na temat słabych i mocnych stron zajęć lub wykaz elementów wymagających poprawy itp.

Reguła 6. Unikaj błędnych założeń, w tym błędnego założenia znawstwa.

Błędne założenia ukryte mogą być zarówno w samym pytaniu kwestionariuszowym, jak i w kafeterii odpowiedzi. Przykładem drugiej możliwości jest przypuszczenie, że prowadzący zajęcia nie mógł być do zajęć całkowicie nieprzygotowany. Założenie takie zawarte zostało w kafeterii odpowiedzi, która pozwala na ocenę stopnia przygotowania prowadzącego do zajęć: „wzorowo”, „bardzo dobrze”, „dobrze” lub „zadowalająco” – nie pozwala jednak na ocenę negatywną. Błędne założenie znawstwa ukryte jest w pytaniu z przykładowego narzędzia do oceny pracy nauczycieli akademickich: „Czy na zajęciach omawiany był najnowszy (z ostatnich lat) dorobek naukowy w tej dziedzinie?” Przypuszczenie, że przeciętny student jest w stanie ocenić stopień zgodności przekazywanego materiału z najnowszym dorobkiem naukowym, jest oczywiście zbyt optymistyczne, w szczególności jeśli weźmiemy pod uwagę stopień specjalizacji poszczególnych dziedzin naukowych. Decydując się na tak sformułowane pytanie, bezwzględnie zawsze powinniśmy umożliwić badanemu wybór kategorii „nie wiem”.

Reguła 7. Zwracaj uwagę na to, żeby kafeteria odpowiedzi była dopasowana (telematycznie i gramatycznie) do zadawanego pytania.

Częstym błędem jest brak gramatycznego dopasowania kafeterii odpowiedzi do pytania kwestionariuszowego. Przykładem takiego „metodologicznego potknięcia” jest pytanie: „Czy prowadzący zajęcia inspirował studentów do samodzielnego myślenia?” połączone z kafeterią odpowiedzi: „wzorowo”, „bardzo dobrze”, „dobrze”, „zadowalająco”.

Reguła 8. Jeśli nie posiadasz merytorycznego uzasadnienia, żeby postępować inaczej, stosuj skale zrównoważone.

Niezrównoważone skale stosuje się najczęściej w momencie wyjątkowo wysokiej skłonności badanych do wybierania odpowiedzi z jednego końca skali (np. w punkcie skrajnie pozytywnym lub negatywnym). Zastosowanie kafeterii: „bardzo dobrze”, „dobrze”, „zadowolająco”, „negatywnie” do pytania kwestionariuszowego: „Jak ocenia Pan/Pani całość zajęć?” związane jest zapewne z założeniem, że większość badanych oceniać będzie pozytywnie zajęcia, w związku z czym skoncentrować należy się na zróżnicowaniu opinii wewnątrz ocen pozytywnych. Równocześnie niezadowolenie z zajęć wyrażone jest w jednej, dodatkowo skrajnej kategorii „negatywnie”. Taki układ proponowanych odpowiedzi może prowadzić do zawyżenia deklarowanych ocen.

Reguła 9. Nie umieszczaj w środku skali szacunkowej kategorii „nie wiem”/ „nie mam zdania”/ „trudno powiedzieć”/ „nie potrafię odpowiedzieć”.

W pierwszej kolejności zastanowić należy się, czy kategoria ta jest w ogóle potrzebna (zazwyczaj generuje kłopotliwe dla analityka braki danych), a jeśli jest konieczna, to umieścić należy ją na końcu skali. Pozostawiając ją w środku skali, narażamy się na zaburzenie struktury i porządku ocen.

Zasady analizy oraz prezentacji materiału empirycznego

Konstruując narzędzie ewaluacji jakości kształcenia w formie kwestionariusza ankiety, musimy zadać sobie pytanie, czego oczekujemy po analizie danych, a tym samym, jaka będzie forma pytań kwestionariuszowych i kafeterii odpowiedzi, czy analizować będziemy pojedyncze wskaźniki, czy raczej będziemy bazować na wskaźnikach syntetycznych w formie skal lub indeksów. Wszystko to prowadzi do przyjęcia już na wstępie co najmniej zarysu planu analizy danych i sposobu prezentacji wyników ewaluacji zajęć dydaktycznych. Sposób analizy i prezentacji rezultatów oceny dydaktyki może wpłynąć na poprawne osza-

cowanie poziomu jakości kształcenia. Z tego też powodu należy zwrócić uwagę na kilka zasad i „pułapek” związanych z analizą oceny zajęć dydaktycznych.

Pierwszym krokiem, który musimy wykonać jeszcze przed rozpoczęciem analizy, jest walidacja danych, czyli sprawdzenie kompletności i poprawności danych ankietowych. Szczególną uwagę należy na to zwrócić wówczas, gdy budujemy indeks sumujący oceny na poszczególnych wskaźnikach w jedną wartość syntetyczną, bowiem w takim przypadku występowanie braków danych w zmiennych wchodzących w skład indeksu może prowadzić do błędnych ocen. Problem ten pojawia się, gdy osoby oceniające unikają odpowiadania na niektóre pytania ankiety ewaluacyjnej, a my wykorzystujemy te pytania w algorytmie obliczeniowym. W takiej sytuacji należy rozważyć, czy w algorytmie obliczeniowym wprowadzić warunek wykluczający braki danych, czy też zastosować jedną z metod imputacji¹⁰. W praktyce istnieje wiele różnych metod uzupełniania braków danych (ang. *missing data imputation*, *missing value imputation*). Najprostszą jest zastąpienie braków średnią ze wszystkich ocen pracowników danej jednostki (wydziału, instytutu) w odniesieniu do danego pytania ankiety ewaluacyjnej. Inne popularne metody imputacji wykorzystują średnią z sąsiednich obserwacji, medianę¹¹ lub interpolację liniową¹². Czasami w celu imputacji braków w ocenach zajęć (pracowników) można zastosować regresję liniową.

Jeśli w ocenie zajęć dydaktycznych wykorzystujemy skale pomiarowe (np. skale Likerta 5- czy 7-stopniowe), a tak w większości przypadków konstruowane są ankiety ewaluacyjne, to schemat analizy danych zakłada porównywanie zajęć lub pracowników względem siebie. W tym celu wykorzystujemy zwykle klasyczne miary położenia (tendencji centralnej), zmienności i asymetrii. Miary tendencji centralnej, tzw. miary przeciętne (średnia arytmetyczna, mediana, modalna), charakteryzują typowy poziom (przeciętne zajęcia, przeciętny pracownik), wokół którego skupiają się wszystkie pozostałe analizowane obserwacje (inne zajęcia, inni pracownicy). Średnia arytmetyczna jest powszechnie stosowana

10 Imputacja to w statystyce sztuczne wstawienie pewnych wartości do tabeli danych. Najczęściej stosuje się ją wtedy, gdy braki danych nie mają losowego charakteru.

11 Mediana zdecydowanie lepiej sprawdza się przy skośności rozkładów danej zmiennej.

12 Interpolacja – metoda numeryczna polegająca na wyznaczaniu w danym przedziale tzw. funkcji interpolacyjnej, która przyjmuje z góry zadane wartości w ustalonych punktach. Stosowana jest wtedy, gdy dysponuje się zazwyczaj skończoną liczbą ocen do określenia zależności między danymi oraz w celu uproszczenia skomplikowanych funkcji arytmetycznych. Interpolacja jest szczególnym przypadkiem metod numerycznych typu aproksymacja. Interpolacja liniowa to przypadek interpolacji za pomocą funkcji liniowej.

w analizie ankiet ewaluacyjnych, bo przez wielu decydentów jest uznawana za intuicyjną miarę, pozwalającą na pozycjonowanie zajęć lub pracowników. Jednakże obliczanie średniej arytmetycznej, o czym nie każdy pamięta, nie jest wskazane, gdy w zbiorze danych występują obserwacje odstające, co bezpośrednio wiąże się z nieproporcjonalnym udziałem obserwacji w tworzeniu sumy. Podobnie, gdy rozkład empiryczny jest silnie asymetryczny, nie należy liczyć średniej arytmetycznej¹³.

Aby skutecznie monitorować i zarządzać jakością kształcenia w szkole wyższej, niezbędnym staje się uwzględnianie w analizie danych rozwiązań z zakresu *benchmarkingu*. W najprostszy sposób można go scharakteryzować jako podejście nakierowane na poprawę własnych działań w oparciu o monitorowanie i porównanie osiągnięć instytucji z osiągnięciami konkurencji bądź innych jednostek, będących elementem szkoły wyższej (por. Woźnicki 2008; Hamalainen 2002). W praktyce wyróżniamy trzy rodzaje *benchmarkingu*: wewnętrzny (szukanie najlepszego wzorca, np. zajęć lub pracownika, i porównywanie się z nim), zewnętrzny – zwany konkurencyjnym (np. porównywanie się z innymi uczelniami) i funkcjonalny (dostosowanie się do wzorca realizującego tę samą lub podobną funkcję, por. Zimniewicz 2009). Przykładowo, na Uniwersytecie Jagiellońskim po zakończeniu oceny zajęć dydaktycznych każdy pracownik może porównać swoje wyniki z wynikami jednostki. W sprawozdaniach analitycznych z oceny zajęć na Uniwersytecie, jakie otrzymuje rektor, uwzględniane są porównania międzywydziałowe, wewnątrzwydziałowe, między pracownikami instytutu lub wydziału (choćby w kontekście typowania najlepszych pracowników na wydziałach do nagród rektorskich), między typami zajęć itp.¹⁴ Wynik takiego benchmarkingu analitycznego może służyć jako podstawa do doskonalenia procesu jakości kształcenia w szkole wyższej na różnych poziomach struktury uczelni¹⁵.

13 Średnia arytmetyczna jest estymatorem zgodnym i nieobciążonym wartości oczekiwanej rozkładu. Jeśli jest to rozkład normalny, to średnia jest również estymatorem efektywnym. Średnia arytmetyczna jest podatna na obserwacje odstające (czyli w tym przypadku wartości zmiennej losowane spoza rozkładu, którego wartość oczekiwaną chcemy estymować, np. pomyłki w danych). W przypadku, gdy jest ich dużo, inne miary tendencji centralnej, takie jak mediana czy średnia obciążona (5% średnia arytmetyczna), mogą dawać lepsze wyniki.

14 Otrzymujemy szereg porównań ocen pracownika w porównaniu ze średnią oceną jednostki, jednostki z innymi jednostkami, ocen jednostki z oceną wydziału, ocen wydziału z oceną globalną uczelni itd.

15 Warto dodać, że systemy *benchmarkingowe* współcześnie coraz częściej skupiają się nie tyle na porównywaniu wartości określonych mierników, co raczej na wymianie informacji o procesie dochodzenia do wysokiej jakości dydaktyki (*benchmarking* procesu).

Aby sensownie porównywać ze sobą zajęcia i pracowników z różnych jednostek (wydziałów lub instytutów) tej samej uczelni (czy też uczelnie między sobą), musimy otrzymywane dane poddać odpowiednim przekształceniom algebraicznym. Wobec danych ilościowych stosowanych w analizach statystycznych formułowane są określone wymogi formalne: warunek symetrii rozkładu, addytywność cech (zwłaszcza gdy są różnoimienne), stałość rozstępu cech (wartości ekstremalnych), jednokierunkowość oddziaływania na zjawisko (Malarska 2005: 39). Spełnienie tych postulatów – jeśli dana cecha ich nie posiada – wymaga zastosowania przekształceń algebraicznych określanych jako normalizacja danych. Normalizacja odnosi się do wstępnej obróbki danych, ujednolicenia cech w celu umożliwienia ich wzajemnego porównywania i dalszej analizy. Powszechnymi rodzajami ujednolicania, tj. normalizacji, które można zastosować do analizy oceny zajęć, są: przekształcenia ilorazowe, standaryzacja, unitaryzacja lub usunięcie ocen odstających. W praktyce wykorzystuje się standaryzację i przekształcenia ilorazowe. Standaryzacja jest rodzajem normalizacji zmiennej losowej, w wyniku której zmienna uzyskuje średnią wartość oczekiwaną zero i wariancję jeden. Najczęściej spotykanym sposobem standaryzacji jest tzw. standaryzacja Z. W opisywanym powyżej przykładzie Uniwersytetu Jagiellońskiego jako formułę normalizacyjną w algorytmie obliczania ocen pracowników uczelni wykorzystuje się przekształcenie ilorazowe oparte na stosunku średniej ważonej z wszystkich zajęć danego pracownika do zakresu ocen (rozstępu skali) dla danego pytania lub baterii pytań. Normalizacja algorytmu analitycznego pozwala uzyskać identyczny dla wszystkich jednostek przedział porównań od 0 do 100 punktów dla zajęć i pracowników, w odniesieniu do semestru, roku akademickiego, jednostki macierzystej, typów zajęć itp. Dzięki temu wynik końcowy jest łatwy w interpretacji dla osób nieobytych ze statystyczną analizą danych.

Na koniec warto dodać, że opisane powyżej metody są tymi, które najczęściej spotyka się w raportach analitycznych z oceny zajęć dydaktycznych, niemniej jednak należy zwrócić uwagę na szereg dodatkowych możliwości w zakresie eksploracji i prezentacji danych z kwestionariuszy ewaluacyjnych. Mamy tu na myśli wspomniane już techniki redukcji przestrzeni danych, takie jak analiza głównych składowych, które umożliwiają identyfikację wymiarowości pojęcia „jakość kształcenia”, a także pozwalają na graficzną prezentację pozycji poszczególnych grup pracowników (np. doktorów, doktorów habilitowanych i profesorów) w wielowymiarowej przestrzeni jakości. W zależności od poziomu pomiaru

zmiennych, którymi dysponujemy, zastosować możemy ponadto szereg technik segmen-
tacyjnych, takich jak choćby analiza skupień.

Rzetelność i trafność studenckich ocen jakości nauczania

Niezależnie od technik badawczych stosowanych do oceny jakości nauczania, podstawo-
wym warunkiem możliwości wykorzystania jej wyników jest pewność, że stanowią one
dobre przybliżenie rzeczywistości, czyli faktycznej jakości prowadzenia zajęć, oraz że są
spójne i stabilne. W metodologii badań kwestie te określa się mianem trafności i rzetelności,
poświęcając im wiele uwagi, bowiem to one decydują o jakości uzyskiwanych wyników,
a co za tym idzie, o jakości i słuszności podejmowanych w oparciu o nie działań i decyzji.
Ponieważ najczęstszą techniką zdobywania informacji na temat jakości nauczania są ba-
dania ankietowe, odniesiemy się tutaj do czynników, które mogą mieć wpływ na wyniki
uzyskiwane za pomocą tych technik.

Trafność

Najprościej rzecz ujmując, trafność odnosi się do tego, czy określone narzędzie mierzy
to, co powinno mierzyć (Babbie 2003:166; Frankfort-Nachmias i Nachmias 2001:180).
Narzędzie służące ocenie jakości nauczania powinno zatem mierzyć faktyczną jakość na-
uczania, być czułe na różnice w jej poziomie i wiernie je oddawać. Jak ujmuje to Wiliam
E. Cashin (1995), trafność w przypadku ankiet studenckich sprowadza się do pytania,
do jakiego stopnia oceny studenckie mierzą jakość i skuteczność nauczania oraz różne
ich wymiary.

Ocena trafności pomiarów jakości nauczania była przedmiotem wielu badań i ekspery-
mentów metodologicznych, cytowanych w artykułach i opracowaniach poświęconych tej
problematyce (por. m.in. Cashin 1995; Erdle, Murray i Rushton 1985; Kulik 2001). W ba-
daniach tych testowano wiele hipotez odnoszących się do różnych wymiarów trafności.
Sprawdzano m.in. czy można zaobserwować związek pomiędzy wynikami uzyskanymi

przez studenta na egzaminie, a tym, jak ocenił on dany kurs (Cohen 1981; Feldman 1989b). Przyjmowano tutaj założenie, że jednym z elementów jakości nauczania jest jego skuteczność, tę zaś można mierzyć poprzez wyniki uzyskane przez studentów na egzaminie¹⁶. W konsekwencji można oczekiwać, że studenci, którzy więcej się nauczyli podczas danego kursu (w praktyce – otrzymali lepsze oceny), będą też lepiej oceniać prowadzącego. Cytowane badania potwierdziły istnienie takiego związku, jednak ich wyniki należy interpretować ostrożnie, ponieważ oceny końcowe studenta zależą od wielu innych czynników, w dużej mierze niezależnych od prowadzącego.

Trafność studenckich ocen kursów potwierdzają też analizy zależności pomiędzy wynikami tych ocen a samooceną prowadzącego oraz ocenami wystawionymi prowadzącemu przez inne osoby: przełożonych, innych nauczycieli akademickich, absolwentów, przeszkolonych obserwatorów zewnętrznych (Feldman 1989a; Overall i Marsh 1980; Braskamp i Ory 1994).

Rzetelność

Jak piszą Chava Frankfor-Nachmias i David Nachmias, „rzetelność dotyczy wielkości błędu związanego z danym narzędziem pomiarowym, tj. błędu, który powstaje w sposób losowy, w kolejnych pomiarach dokonywanych za pomocą tego samego narzędzia pomiarowego” (2001:185). W literaturze poświęconej mierzeniu jakości kształcenia przez rzetelność rozumie się zazwyczaj spójność i stabilność ocen (por. Cashin 1995), a w ramach analizy ankiet studenckich rzetelność odnosi się najczęściej do spójności ocen pomiędzy poszczególnymi oceniającymi (np. czy wśród uczestników jednego kursu, prowadzonego przez jednego uczącego istnieje zbieżność ocen dla poszczególnych pozycji ankiety).

Na rzetelność studenckich ocen zajęć wpływa wiele czynników. Kluczowe znaczenie odgrywa sama technika badawcza oraz wykorzystywane narzędzie badawcze, w tym przede wszystkim jasność, jednoznaczność i zrozumiałość pytań, o czym pisaliśmy w poprzedniej części tego opracowania. Niezmiernie istotną kwestią jest też liczba i sposób doboru studentów oceniających dane zajęcia: generalnie im większy odsetek oceniających, tym bardziej rzetelne wyniki. W literaturze poświęconej ocenie jakości nauczania spotkać można

16 W eksperymentach wykorzystywano egzaminy przygotowywane nie przez samego prowadzącego, lecz przez osoby zewnętrzne, aby zminimalizować wpływ różnicowania stopnia trudności egzaminów.

konkretne zalecenia dotyczące tego, jak duży musi być poziom realizacji próby dla danej wielkości grupy studentów, aby oceny mogły zostać uznane za reprezentatywne i rzetelne (por. Thall i Franklin 1991). Zalecenia prezentujemy w poniższej tabeli:

Tabela 1. Zalecany poziom realizacji próby w studenckich ocenach zajęć

WIELKOŚĆ GRUPY	ZALECANY POZIOM REALIZACJI PRÓBY
5–20	80%
21–30	75%
31–50	66%–75%
51–100	60%–75%
Ponad 100	50%–75%

Źródło: opracowanie własne na podstawie M. Theall, J.I. Franklin (1991)

Istotny jest jednak nie tylko sam poziom realizacji próby, czyli odsetek uczestników danych zajęć, którzy wypełnili ankietę, ale i reprezentatywność próby, którą możemy tutaj rozumieć przede wszystkim jako brak występowania autoselekcji podczas wypełniania kwestionariusza. Autoselekcja może bowiem w tym przypadku stanowić źródło systematycznych obciążeń uzyskanych ocen. Bardziej chętni do wypełnienia ankiety mogą być studenci silnie zmotywowani do wyrażenia swojej opinii, czyli na przykład bardzo z zajęć zadowoleni bądź też bardzo niezadowoleni.

Problem ten w dużym stopniu dotyczy wspomnianych przez nas wcześniej ankiet internetowych (Computer Aided Web Interview, w skrócie CAWI), stosowanych coraz powszechniej w badaniach opinii studentów na temat zajęć. O wzrastającej popularności tej techniki decyduje przede wszystkim łatwość jej zastosowania (zwłaszcza w sytuacji, gdy uczelnia dysponuje elektronicznym systemem obsługi studentów), niski koszt zbierania danych, szybkość zbierania danych, redukcja błędów przy wprowadzaniu danych i ogólna łatwość zarządzania takimi badaniami. Atuty te doskonale streszcza Jadwiga Przewłocka, pisząc, że technika „opierając się na nowych technologiach i umożliwiając atrakcyjną z punktu widzenia respondenta formę kontaktu, pozwala na szybkie uzyskiwanie dużych prób stosunkowo niskim kosztem” (2009:99). Jak w każdym badaniach internetowych, problemem pozostają jednak wspomniane wyżej: niski poziom realizacji próby oraz autoselekcja badanych.

W badaniach studenckich ocen zajęć realizowanych za pomocą ankiety CAWI mamy zazwyczaj do czynienia z próbami wyczerpującymi, czy też inaczej, z badaniami realizowanymi na całych populacjach: zaproszenie do wzięcia udziału w badaniu kierowane jest do wszystkich uczestników danego kursu. Uczestnicy ci, zachęcani na różne sposoby, m.in. przez bannery pojawiające się na stronie obsługi studentów, plakaty rozwieszane na uczelniach, zachęty prowadzących zajęcia, artykuły i apele w prasie uczelnianej, mogą wypełnić przypisaną do danych zajęć ankietę. Zazwyczaj takich ankiet do wypełniania mają kilka: liczba ankiet odpowiada liczbie kursów, w jakich wzięli udział i formom tych kursów – zazwyczaj oddzielnie wypełnia się ankietę dla wykładów, ćwiczeń, laboratoriów czy praktyk¹⁷. Tak założony schemat realizacji badania sprawia, że wszystkie osoby mają jednakowe szanse na wzięcie udziału w badaniach, zakłada się bowiem, że każdy student uczęszczający na zajęcia ma dostęp do Internetu, mając tym samym możliwość wypełnienia ankiety. Jak jednakże pisze Przewłocka:

odpowiedni dobór próby nie zapewni jednak wysokiej jakości próby, jeśli wylosowane do próby jednostki nie wezmą udziału w badaniu. Jest to tzw. problem braku odpowiedzi (*non-response error*), którego skala zależy z jednej strony od stopnia realizacji próby (*response rate*), z drugiej – od różnic między osobami, które wypełniły ankietę i tymi, które jej nie wypełniły. (2009:104)

Niski poziom realizacji próby, będący jednym z najpoważniejszych problemów badań CAWI może też silnie wpływać na reprezentatywność wyników uzyskanych dzięki ankietom studenckim. Niechęć do wypełniania ankiet w przypadku oceny zajęć może być dodatkowo wzmacniana przez swoisty brak kultury ewaluacji na polskich uczelniach. Ów brak kultury ewaluacji, związany z nowością działań z tego zakresu i często z brakiem przekonania o ich znaczeniu, widoczny jest zarówno wśród studentów, jak i wśród prowadzących. Jak można przypuszczać, istnienie tego zjawiska zdecydowanie nie sprzyja gotowości do uczestnictwa w badaniach oceniających jakość nauczania.

Naturalnie, podejmowanie działań zachęcających do wzięcia udziału w ocenie zajęć, w tym podkreślanie anonimowości tych badań, często niepokojącej studentów i kwestionowanej przez nich, może wpłynąć na poziom realizacji próby, zwiększając tym samym prawomoc-

17 Sytuacja ta może się jednak w praktyce komplikować, gdy jedno zajęcia, np. wykład z danego przedmiotu, prowadzone są przez kilku wykładowców.

ność generalizacji. Problemem pozostaje jednak wciąż możliwość występowania różnic pomiędzy tymi, którzy ankietę wypełnili i tymi, którzy jej nie wypełnili.

Stabilność z kolei jest związana ze spójnością ocen danego kursu, ewaluowanego przez różnych oceniających w różnych momentach czasu. Jak pokazują badania (por. np. Braskamp i Ory 1994; Centra 1993; Overall i Marsh 1980), oceny tego samego prowadzącego są raczej podobne w różnych punktach czasowych. **Możliwość uogólniania** określa związek pomiędzy pojedynczymi ocenami danego prowadzącego a jego rzeczywistymi umiejętnościami jako wykładowcy. Przeprowadzono wiele badań dotyczących tej problematyki (por. np. Marsh 1982), a wyniki wskazują, że to osoba prowadzącego i sposób prowadzenia zajęć w większym stopniu wpływa na oceny studenckie niż sam kurs (jego zakres tematyczny, przedmiot itp.).

Możliwe źródła obciążenia danych

Obciążać wyniki ocen danego prowadzącego, czy oceny danego kursu mogą te wszystkie czynniki, które nie są zależne od prowadzącego czy też nie wynikają ze specyfiki kursu, a wpływają na efektywność uczenia lub zadowolenie studentów z uczestnictwa w zajęciach. Do takich czynników należą np. wielkość grup ćwiczeniowych, warunki, w jakich odbywają się zajęcia, ale też zainteresowanie studentów tematyką zajęć.

Wskazuje się, że podczas analizy wyników ankiet studenckich, a zwłaszcza przy wykorzystaniu ich w polityce kadrowej, warto kontrolować te czynniki, które mogą obciążać pomiar. Prowadzone badania pozwoliły wyróżnić wiele takich czynników. Najważniejsze z nich przedstawia poniższa tablica:

Tablica 2. Czynniki wpływające na studenckie oceny zajęć

- **Motywacja studentów do uczestnictwa w kursie**, zainteresowanie problematyką. Wyżej oceniane są zazwyczaj kursy, które studenci wybierają dobrowolnie lub gdy uczestniczą w kursach związanych z ich zainteresowaniami.
- **Oczekiwana ocena**: wyżej oceniane przez studentów są kursy, z których otrzymują lepsze oceny. Wyniki badań nie są pod tym względem jednoznaczne, podobnie jak wyjaśnienia przyczyn tego zjawiska. W jednym z przytaczanych wyjaśnień wskazuje się, że ci prowadzący, którzy wystawiają studentom lepsze oceny od tych, na jakie zasłużyli, są także przez studentów lepiej oceniani, niż na to zasłużyli.
- **Poziom kursu, miejsce w cyklu studiowania**: różnie oceniane są kursy podstawowe i bardziej zaawansowane.
- **Dyscyplina naukowa**: badania wskazują, że oceny kursów związane są z kierunkiem studiów. Najwyżej oceniane są kursy z dziedziny nauk humanistycznych, następnie z dziedziny nauk społecznych, najniżej z dziedziny nauk ścisłych. Oznacza to, że ocen tych kursów nie można porównywać.
- **Stopień trudności kursu**: wbrew potocznym opiniom, wiele badań wskazuje jednak, że wyżej oceniane są te kursy, które wymagają większego nakładu pracy, są trudniejsze.
- **Brak anonimowości**: wpływa na wyższe oceny kursu/prowadzącego.
- **Obecność prowadzącego zajęcia** podczas oceniania: również wpływa na wyższe oceny kursu.

Źródło: opracowanie własne na podstawie Cashin (1995)

Obawy związane z oceną zajęć dydaktycznych

Ostatnią rzeczą, na którą należy zwrócić uwagę, są obawy nauczycieli akademickich wobec wyników oceny zajęć dydaktycznych, sposobu ich wykorzystania oraz prezentacji i upubliczniania w społeczności akademickiej danej szkoły wyższej. Do najczęstszych obaw związanych z ewaluacją zajęć dydaktycznych można zaliczyć: obawę przed byciem

ocenianym oraz obawy związane z gloryfikowaniem ewaluacji jako źródła pozyskiwania informacji zwrotnej, szczególnie, gdy jest to jedyna forma gromadzenia informacji na temat jakości prowadzonych zajęć w szkole wyższej. Obawy budzi także stosunek studentów do wypełniania ankiet ewaluacyjnych, co rodzi pytanie o rzetelną informację zwrotną i obiektywność ocen studenckich. Oceniani mają poczucie, że studenci traktują ankiety ewaluacyjne jako sposób karania i nagradzania pracowników naukowo-dydaktycznych. Dochodzi do tego przekonanie o tym, że ankiety wypełniają albo studenci zadowoleni, albo niezadowoleni (co może prowadzić do obciążenia wyników ocen). W takiej sytuacji wykorzystywana w pozycjonowaniu zajęć lub pracowników średnia arytmetyczna jest wypadkową ocen skrajnych.

Często artykułowane są także obawy związane z niedoskonałością narzędzi ewaluacyjnych m.in. wielością pytań (szczególnie w przypadku ankiet elektronicznych, gdy student ma do wypełnienia przeciętnie po kilka ankiet w semestrze) i niejednoznacznością pytań (gdy student ma problem z udzieleniem odpowiedzi zgodnej z własnymi przekonaniem i/lub stanem faktycznym). Należy jednak zdać sobie sprawę, iż kwestionariusz ankiety ewaluacyjnej wymaga wielu testów i czasu zanim stanie się rzeczywiście trafnym i rzetelnym narzędziem oceny jakości kształcenia.

Ostatnią kategorią są obawy dotyczące sposobu prezentacji i wykorzystania wyników ewaluacji. Szczególnie w przypadku ankiet elektronicznych podkreślane są dwie kwestie: wiarygodność oceny (jaki powinien być odsetek oceniających/ile ankiet trzeba wypełnić, aby wyniki uznać za miarodajne) oraz późniejsze konsekwencje dla pracownika (problem transmisji oceny z ewaluacji na ocenę przełożonego, szczególnie w sytuacji, gdy nie jest ona adekwatna do ewaluowanej rzeczywistości). Niezależnie od tego występuje obawa, że wynik ewaluacji może stanowić podstawę do rozwiązania stosunku pracy z nauczycielem akademickim lub być przyczyną braku propozycji prowadzenia zajęć dydaktycznych.

W przypadku studentów obawy w zakresie oceny zajęć dydaktycznych są głównie związane z brakiem lub pozorami zachowania anonimowości (co może przekładać się na negatywne konsekwencje wobec studentów), zachowaniem poufności danych, niedoskonałością narzędzia ewaluacyjnego i sposobem wykorzystania wyników ewaluacji. Do tych obaw dochodzi także problem upubliczniania wyników ewaluacji wśród studentów i pracowników danej uczelni. Należy bowiem pamiętać, że istnieje wyraźny związek między powszechnością

wyników a frekwencją oceniających. Raporty z oceny winny być publikowane na stronie internetowej uczelni oraz na stronach internetowych jednostek dydaktycznych, tak aby oceniający mieli wgląd w wyniki. A to dla wielu decydentów stanowi poważny problem.

Jak widać, zarówno studenci jak i pracownicy, wśród swoich obaw odnośnie oceny zajęć dydaktycznych podkreślają dwie kluczowe kwestie: posiadanie dobrego, tzn. trafnego i rzetelnego, narzędzia ewaluacyjnego oraz sposób wykorzystania, prezentacji i upubliczniania wyników. Dlatego wszelka ocena – zwłaszcza jeśli ma ona mieć praktyczne konsekwencje (a ocena jakości kształcenia ma je z pewnością) – powinna spełniać cztery postulaty: **prawdziwości, dokładności, wszechstronności i uczciwości**. Warto więc zadbać o te postulaty, aby nie tylko doskonalić proces zbierania i analizy oceny zajęć dydaktycznych w każdej szkole wyższej, ale mieć realny wpływ na zarządzanie jakością kształcenia.

Zakończenie

Niniejszy tekst należy traktować jako zestawienie refleksji na temat metodologii konstruowania narzędzia i analizy wyników ankiet oceniających pracę nauczycieli akademickich. Refleksje te dotyczyły kwestii związanych z trafnością i rzetelnością pomiaru, standardami konstrukcji pytań kwestionariuszowych czy sposobem prezentacji materiału empirycznego. Świadomi jesteśmy, że tematy poruszone przez nas w artykule nie wyczerpują problematyki, a raczej mają na celu pobudzenie do wspólnej refleksji nad metodologicznym aspektem ewaluacji praktyki dydaktycznej. Tak jak świadomi jesteśmy, na co zwracaliśmy już uwagę, że ewaluacja jakości kształcenia w szkole wyższej nie sprowadza się do eksploracji ankiet ewaluacyjnych, a ankiety te są jedynie drobnym, dodatkowo narażonym na duży subiektywizm oceniających, elementem pomiaru szeroko rozumianej jakości kształcenia. Ponadto ewaluacja jakości kształcenia nie jest ewaluacją samego prowadzącego kurs, ale raczej powinna skupiać się na procesie, w którym uczestniczą nauczyciel akademicki i student. Proces ten dodatkowo odbywa się w określonej przestrzeni (sale wykładowe, laboratoria, wyposażenie), która może mieć wpływ na samą jakość kształcenia. Każdy z tych elementów poddany powinien zostać dogłębnej ocenie.

Współczesne definicje jakości kształcenia raczej rozszerzają, a nie zawężają pojęcie, sytuując je zazwyczaj na tle terminów, takich jak: skuteczność i efektywność kształcenia, osiągnięcia szkolne czy produktywność nauczania. Rodzi to wymóg zastosowania wielu metod, podejść, zespolenia koncepcji i opracowania nowych narzędzi pomiarowych. Implikuje to ponadto głębsze spojrzenie w sam proces funkcjonowania szkoły wyższej, który powinien być analizowany z uwzględnieniem różnorodnych źródeł i typów danych.

Przykładowo, pamiętać należy, że jakość kształcenia to nie tylko zadowolenie ze studiowania, ale też zdobyta przez studentów wiedza/umiejętności i ich użyteczność, ich sukces zawodowy i społeczny, a także wizerunek tej jakości ukształtowany w świadomości kandydatów na studia. Tym samym informacje o jakości kształcenia ukryte są nie tylko w ankietach ewaluacyjnych, ale też: w wynikach badania losów absolwentów, w danych sprawozdawczych szkół wyższych (np. sprawozdania składane do GUS, informacje o osiągnięciach nauczycieli akademickich raportowane w ramach ankiety oceny jednostki naukowej przygotowywanej dla OPI, dane z działów zajmujących się rekrutacją studentów), a nawet w ocenach i wyborach kandydatów na studia.

Sprawne wykorzystanie narzędzi badawczych wymaga określenia wspólnych standardów ich konstrukcji, czy to na poziomie uczelnianym, czy nawet ogólnopolskim. Nie oznacza to oczywiście, że szkoły wyższe zmuszone są do stosowania tego samego „szablonu” podczas ewaluacji jakości kształcenia, a wszelka innowacyjność powinna być w nich ograniczana. Sugeruje się raczej wspólne wypracowanie kilku wskaźników, które pozwolą na porównywanie ocen pomiędzy uczelniami.

Na koniec warto dodać, że wzrost świadomości znaczenia badań prowokuje do dyskusji na temat poprawności wnioskowania w procesie oceny pracy nauczycieli akademickich. Miejmy nadzieję, że wzrost ten doprowadzi do rozwoju metodologii ewaluacji zajęć dydaktycznych i do opracowania wspólnych reguł konstruowania narzędzi i analizy materiału empirycznego.

Bibliografia:

Babbie, Earl (2003), *Badania społeczne w praktyce*, Warszawa: PWN.

Braskamp, Larry A., Ory, John C. (1994), *Assessing Faculty Work: Enhancing Individual and Institutional Performance*, San Francisco: Jossey-Bass.

Brzezińska, Anna, Brzeziński, Jerzy (2004), *Skale szacunkowe w badaniach diagnostycznych*, [w:] Jerzy Brzeziński (red.), *Metodologia badań psychologicznych. Wybór tekstów*, Warszawa: PWN, s. 232–307.

Cashin, William E. (1995), *Student Ratings of Teaching: The Research Revisited*, „IDEA Paper”, 32, Manhattan, KS: Center for Faculty Evaluation and Development, Division of Continuing Education, Kansas State University.

Cave, Martin, Hanney, Stephen, Henkel, Mary, Kogan, Maurice (2006), *The Use of Performance Indicators in Higher Education. The Challenge of the Quality Movement*, „Higher Education Policy Series”, 34.

Centra, John A. (1993), *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*, San Francisco: Jossey-Bass.

Cohen, Peter A. (1981), *Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies*, „Review of Educational Research”, 51, s. 281–309.

Erdle, Steve, Murray, Harry G., and Rushton, Philippe J. (1985), *Personality, Classroom, Behavior, and College Teaching Effectiveness: A Path Analysis*, „Journal of Educational Psychology”, 77, s. 394–407.

Feldman, Kenneth A. (1989), *The Association Between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data from Multisection Validity Studies*, „Research in Higher Education”, 30, s. 583–645.

Gaither, Gerald, Nedwek, Brian, Neal, John (1994), *Measuring Up. The Promises and Pitfalls of Performance Indicators in Higher Education*, „ASHE-ERIC Higher Education Report”, 23(5).

Gates, Susan, Augustine, Catherine, Benjamin, Roger, Bikson, Tora, Kaganoff, Tessa, Levy, Dina, Moini, Joy, Zimmer, ROn (2002), *Ensuring Quality and Productivity in Higher Education. An Analysis of Assessment Practices*, „ASHE-ERIC Higher Education Report”, 29(1).

Hamalainen, Kauko (2002), *Benchmarking in the Improvement of Higher Education*, ENQA Workshop Reports 2, Helsinki.

Kulik, J.A. (2001), *Student Ratings: Validity, Utility, and Controversy*, „New Directions for Institutional Research”, 109, s. 9–25.

Marsh, Herbert W. (1982), *The Use of Path Analysis to Estimate Teacher and Course Effects in Student Rating's of Instructional Effectiveness*, „Applied Psychological Measurement”, 6, s. 47–59.

Nachmias-Frankfort, Chava, Nachmias, David (2001), *Metody badawcze w naukach społecznych*, Poznań: Zysk i S-ka.

Oppenheim, Abraham (1992), *Kwestionariusze, wywiady, pomiary postaw*, Poznań: Zysk i S-ka.

Overall, Jesse U., Marsh Herbert W. (1980), *Students' Evaluations of Instruction: A Longitudinal Study of Their Stability*, „Journal of Educational Psychology”, 72, s. 321–325.

Przewłocka, Jadwiga (2009), *CAWI – specyfika, wykorzystanie, perspektywy rozwoju*, [w:] Agnieszka Haber, Maciej Szałaj (red.), *Ewaluacja wobec wyzwań stojących przed sektorem finansów publicznych*, Warszawa: PARP.

Sztabiński, Paweł, Sawiński, Zbigniew, Sztabiński, Franciszek (2005), *Fieldwork jest sztuką*, Warszawa: Wydawnictwo IFiS PAN.

Theall, Michael, Franklin, Jennifer (2001), *Looking for Bias in all the Wrong Places: a Search of Truth or a Witch Hunt in Student Ratings of Instruction?*, „New Directions for Institutional Research”, 109, s. 45–56.

Theall, Michael, Franklin, Jennifer (red.)(1990), *Student Ratings of Instruction: Issues for Improving Practice. New Directions for Teaching and Learning*, 43, San Francisco: Jossey-Bass.

Theall, Michael, Franklin, Jennifer (1991), *Using Student Ratings for Teaching Improvement*, [w:] Michael Theall, Jennifer I. Franklin (red.), *Effective Practices for Improving Teaching. New Directions for Teaching and Learning*, 48, San Francisco: Jossey Bass.

Woźnicki, Jerzy (red.)(2008), *Benchmarking w systemie szkolnictwa wyższego*, Warszawa: FNP.

Yorke, Mantz (1996), *Indicators of Programme Quality: a Project Report Prepared for HEQC*, London: Higher Education Quality Council.